Identification of epigenomic patterns to annotate regulatory elements in the human genome.

Utrecht University

Jeske van Riel (3220257)
J.J.G.vanriel@students.uu.nl
Utrecht University
Faculty of Beta Sciences
Department of Biomedical Sciences
Cancer Genomics and Developmental Biology

Supervised by Menno Creyghton, PhD (Hubrecht Institute)

Abstract

Recently, studies on transcriptional regulation have turned from single-gene to genome-wide research. This transgression provides new possibilities towards the identification of transcriptional regulation pathways, and the identification of new transcriptional regulatory elements. Recent technological advances have enabled the identification of histone modifications. These have been mapped to specific chromatin structures and transcriptional regulatory elements, such as promoters or enhancers. Specific patterns of histone modifications, regulatory proteins or chromatin structural properties have been discovered. These patterns were shown to demarcate specific regulatory elements, and are being increasingly used to identify novel elements.

The resulting data from these studies show an intricate interplay between posttranslational histone modifications, chromatin structure, and genome function throughout different cellular conditions. It is considered that histone modification patterns enable prediction of functional regulatory elements, gene expression and splicing patterns. Increased knowledge on these patterns has facilitated research on different genomic expression profiles between healthy and diseased cells in many disorders and diseases. Many studies on cancer have indicated an important role for epigenetic regulation during cancer development, commending the important role of the search for epigenetic patterns and their potential function.

Clear histone modification patterns have been identified for coding regions, promoters and enhancers. The negative regulatory elements, insulators and silencers, however are greatly underrepresented in these studies, and no definite histone modification patterns have been described so far. In order to get a full view of the intricate regulation of transcription, it seems vital to improve our understanding of histone modification patterns on all regulatory elements. In this review, an overview of histone modification patterns on regulatory elements, and their importance for transcription regulation is given. Special attention is directed to the patterns on silencers and in silenced genomic regions.

Table of Contents

Chapter 1 Regulation of gene expression

The Human Genome Project has shown that there are approximately 20.000 protein-coding genes in the human genome, all of which have a specific location and function to be carried out.[1] All cells in the human body require the exact spatial and temporal expression of genes in order to function accurately. For the correct regulation of transcription, the genome contains specific transcriptional regulatory elements. These elements are often located in non-coding regions of the DNA and so far, several have been identified.[2] In order to fully understand the complexity of the human genome, all regulatory elements have to be identified. Their location, interactions with other parts of the genome and expression-regulating proteins, and other specific characteristics, will greatly improve our knowledge of the regulation of transcription. This will improve our understanding of inter- and extracellular pathways, which can aid in the timely identification of diseases, or provide new means for remedies. Though we possess most of the human genome sequence, identification of these regulatory elements has been tedious. To improve this process, identification methods other than comparative studies are required. One such method has been sought in patterns of posttranslational histone modifications.[3] These modifications have been found on all histone proteins present in cells and are thought to form specific patterns correlating with genomic regulatory elements. Great effort has been put in the identification of these patterns, their function, and possible correlation with regulatory elements. For some regulatory elements, such as enhancers or promoters, histone modification patterns have been discovered.[4] These patterns are being used for various applications, such as identification of new regulatory elements in the human genome. Unfortunately, patterns have not been identified for all regulatory elements. Silencers for example are heavily under-researched and much remains to be discovered. In this work the annotation of histone modifications patterns to regulatory elements will be discussed. First, the regulation of transcription and its different regulatory elements will be depicted. Next, histone modification identification techniques, histone modification patterns and their corresponding regulatory elements will be discussed. In chapter 3, silenced genomic regions, silencers, and their potential histone modification patterns will be described in detail. Last, the functional applications of these histone modification patterns will be discussed.

The genome of eukaryotes is structured into chromatin

Regulation of transcription in eukaryotes happens, for the first part, in the nucleus, where protein-coding genes are translated into mRNA. The eukaryotic genome is tightly packed into chromatin, a nucleoprotein complex inside the nucleus, which is composed of nucleosomes and DNA. A nucleosome consists of 147 bp of DNA wrapped around a histone octamer, which generally contains two copies of each of the core histones H2A, H2B, H3 and H4. These nucleosomes are separated by approximately 16 bp of linker DNA, which associates with one copy of histone H1. Histone H1 is present to keep the DNA in place, and aids in the formation of the "beads-on-a-string" chromatin structure.[5] In actively transcribed regions, the DNA often is more loosely wrapped around its histone octamer. This increases accessibility for the transcriptional machinery. These active regions are referred to as euchromatin. Inactive DNA regions consist of heterochromatin, which is packaged tightly to prevent binding of the transcriptional machinery. This difference is often referred to as the secondary structure of chromatin. When packaged tightly it is called the "30-nm fiber", referring to the width of the complex. When loosely packaged chromatin is also referred to as "10-nm fiber".[6] Throughout the genome, some variants of the core histones are incorporated in nucleosomes (table 1). These can be H2A.Z, H2A.X, Macro H2A, CENP-A or H3.3. They diverge in their sequence from the original core histone in varying degrees, and are incorporated to execute diverse functions. H2A.Z for example, replaces H2A in the nucleosome and is often found in the promoters of actively transcribed genes. It is considered to prevent the spreading of heterochromatin to these regions.[7]

Each histone protein contains a globular domain, important for histone-histone interactions, and a highly dynamic amino terminal tail extending outwards from the nucleosome. These extruding tails are approximately 20-35 residues in length, and rich in basic amino acids.[6] They are prone to

posttranslational modifications (PTMs), which have been the interest of many studies. PTMs found on histones include acetylation and methylation of arginines; phosphorylation of serines and threonines; and acetylation, methylation, ubiquitination, sumolyation and biotinylation of lysines.[6,10]

Table 1. Different variants of core histones and their functions in transcriptional regulation.[8,9]

Histone Variant	Similarity with core histone	Functions
H2A.X	96%	DNA damage repair; the Serine in the SQEY motif in mammalian H2A.X is phosphorylated in response to double strand DNA breaks.
H2A.Z	60-65%	Transcriptional regulation, DNA damage and chromosome stability.
Macro H2A	64%	Enriched on inactive mammalian X chromosome.
CENP-A	Highly diverged	Centromere-specific, specialized H3 variant for attachment of kinetochores.
H3.3	97%	Replaces H3.1 at transcriptionally active regions, varies only at 5 amino acids.

Addition of one, two or three methyl groups to lysines is probably the most common modification present on histone tails.[10] Methylation does not affect the charge of the lysine and adds a minimal amount of atoms, resulting in only a slight effect on its chemistry. This is also true for the methylation of arginines, although these are not trimethylated.[11] These modifications therefore do not alter the interaction with the DNA. They however do provide recognition sites for proteins containing specific domains (Table 2). Chromo and PHD domains are able to recognize lysine methylation and differentiate between mono-, di- or trimethylation. Tudor domains can recognize methylated arginines.[11] In contrast, addition of an acetyl group does have a major effect on the positive charge of lysines. They become more neutrally charged, causing an increased dissociation of the DNA from the nucleosome.[11] Acetyltransferases, the enzymes responsible for acetylation, are often not entirely accurate and acetylate multiple lysines at once.[11] High levels of acetylation in a genomic region can cause the loosely packaging of euchromatin. Phosphorylation of serines or threonines also adds a negatively charged group to the amino acid. It has however yet to be shown what implications this has on the histone or DNA structure.[11] Throughout the genome many PTMs can be found on histones, having different functions involving transcriptional regulation. Many PTMs have been shown to form specific patterns coinciding with genomic regulatory elements.

Transcription of genes is regulated by different sets of transcription factors and regulatory elements

Genomic regulatory elements are DNA sequences involved with the regulation of expression, through association with transcription factors. Genes usually contain a proximal regulatory element, the promoter, and are influenced by several distal regulatory elements, such as enhancers and silencers. The promoter can be found directly upstream from a gene, and contains the transcription start site (TSS). Promoter are composed of a core promoter and a proximal promoter, which consist of several regulatory elements.[14]

Expression of genes is regulated at several points during transcription, such as at the formation of the preinitiation complex, transcription initiation and elongation.[12] In eukaryotes, transcription is performed by the RNA Polymerase II protein complex. Transcription involves many different transcription factors, of which the general transcription factors (GTFs) are best described. In this group, RNA Polymerase II, TFIIA, TFIIB, TFIID, and other important factors of the classic transcription machinery can be found.[12] During transcription initiation these GTFs assemble in a specific order on the DNA to form the transcription preinitiation complex (PIC), which directs RNA Polymerase II to the transcription start site (TSS). After formation of the PIC, transcription initiation follows an ordered series of events to form a functional RNA Polymerase II elongation complex.[12]

The core promoter serves as a docking system for the transcription machinery and PIC assembly.[16] The best known example of a regulatory element within the core promoter is the TATA box, a binding site for TFIID during PIC formation. Other well-known elements include the Initiator Element (Inr), the Downstream Core Element (DCE), the Downstream Promoter Element (DPE), and the TFIIB-Recognition

Element (BRE). Although all these elements are present in promoters, most promoters do not contain all of them. The TATA box for example, is only present in approximately 24% of all human promoters.[16,17]
The combination of different regulatory elements in promoters enables the relatively few DNA-binding transcription factors to regulate all protein-coding genes. Estimates on the total amount of transcription factors in humans differ from 1500 to 3000.[15] Many of these have been mentioned throughout literature, however only less than 100 have been experimentally proven so far.[15] Regulating in total over 20,000 protein-coding genes, transcription factors rely greatly on combinatorial regulation to remain specific. The different regulatory elements on promoters confer combinatorial control of gene regulation, causing an exponential increase in the number of unique expression patterns. Despite the fact that many of these combinations are uninformative, combinatorial control still increases the number of possible combinations and specificity.[16]

In order to stimulate transcription, a second class of transcription factors is required. These are activators and co-activators, and they greatly increase the level of transcription. They can form hetero- or homodimers and are specific in their binding to DNA sequences, working synergistically to increase transcription activity. These transcription factors are known to promote any step of the transcription process. They can assist PIC formation, by recruiting some of the proteins involved in this complex. Some activators can increase the rate of promoter escape or influence the activity of Polymerase II, while others have been shown to facilitate re-initiation.[13] Much remains to be discovered about these mechanisms of stimulation, and they are therefore under intense research.[14] Transcription factors have also been proposed to increase transcription by modification of the chromatin structure. This can happen through ATP-dependent remodeling complexes or histone-modifying complexes, to enhance availability of the DNA to the transcription complex. Activators can also function as docking sites for additional factors involved in the enhancement of transcription.[14] The combination of these different transcription factors facilitates the complicated regulation of genomic transcription.
The proximal promoter is the region immediately upstream from the core promoter.[14] It contains binding sites for activators. Approximately 50-60% of all vertebrate promoters are found near CpG-dinucleotide islands, a short stretch of DNA with a relatively high guanine and cytosine content. CpG-islands near promoters are generally unmethylated. These unmethylated CpG-islands are considered a reliable source of promoter region localization.[18] Unfortunately not all promoters are associated with unmethylated CpG-islands, indicating the requirement for different identification methods.

Conclusively, regulation of transcription is initiated at gene promoters and influenced by a diverse set of transcription factors, and several distal regulatory elements.

Distal regulatory elements regulate expression of genes from further away
Aside from the promoter, distal regulatory elements are also involved in the regulation of transcription (Figure 1). While transcription initiation happens at the promoter, distal regulatory elements enhance or diminish the expression of genes. Thus far, several distal regulatory elements have been identified. These are enhancers, insulators, silencers, and locus control regions.

Enhancer
From all distal regulatory elements, the enhancer has been studied most intensively. Enhancers were first identified in the genome of the SV40 virus, which was shown to be able to increase transcription of their hosts genome.[19] Enhancers typically exist of closely positioned transcription factor binding sites (TFBSs), which collaborate to enhance transcription. They have been shown to regulate transcription in a spatial- and temporal-dependent manner, but will function regardless of their distance and orientation relative to the gene its acting upon.[20] Enhancers are able to interact with several genes, and different genes can be influenced by multiple enhancers. This diversity of enhancer-gene interactions aids in the spatial- and temporal-dependent regulation of genes. When genes require

activation through multiple enhancers, the absence of one of these enhancers can prevent its transcription.

Enhancers are often compared to the proximal promoter of genes, because they can bind the same transcription factors, and are both able to enhance gene expression. There are however several difference between them. First, they differ in their position. While the proximal promoter is always directly upstream from the expressed gene, enhancers can be hundreds of kilobases up- or downstream from the actual gene.[20] Distances of up to 1 Mb between enhancers and the TSS of the gene they interact with have been discovered.[21] Enhancers can also be found on different chromosomes.[22] Second, the promoter contains the transcription start site for the gene it's regulating, while the enhancer can interact with the promoter to increase transcription initiated at this promoter. In order for the promoter and enhancer to communicate, it is considered that the DNA loops. This brings both elements in close proximity, enabling the enhancer to perform its function.[20] Third, the functionality of the promoter is dependent on its position and orientation,[14] while enhancers can often still function when inverted or placed elsewhere.[20]

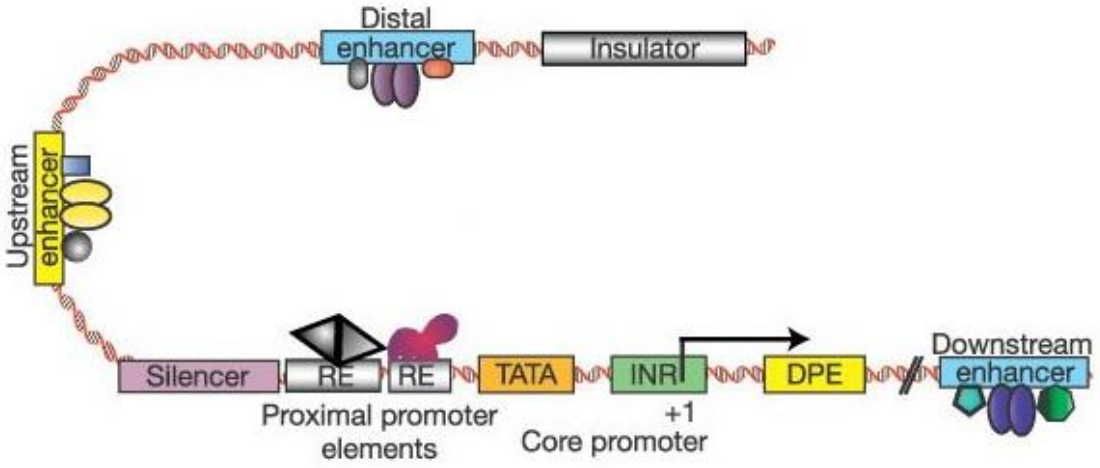

Figure 1. Gene expression is regulated by the promoter and distal regulatory elements.

Directly upstream of the gene is the promoter, which exists of the core promoter and proximal promoter elements. Distal regulatory elements either increase (enhancers) or decrease (silencers) gene expression. Both enhancers and silencers are can be located in close proximity of the gene or further up- or downstream. Insulators are boundary elements, preventing the influence of regulatory elements to spread to nearby genes.

Silencers

The antagonists of enhancers are the silencers. Silencers are distal regulatory elements able to decrease or repress transcriptional activity. They can be found as an independent distal element, but can also be part of the proximal promoter, or even of a distal enhancer.[23] Silencers contain binding sites for repressors, negative regulatory transcription factors. Repressors function synergistically, and sometimes require the assistance of corepressors.[24] Repressors have been found to repress PIC formation. They can also function by blocking the binding of nearby activators, or by competing for the same binding site. Repressors can also affect chromatin structure, creating a repressive environment to prevent activating transcription factors from reaching their required destination.[24] Because most genes do not require expression at any given time, the repressive state of most genes is considered default, and they are only activated when specifically needed.[23] Most genes are therefore located in heterochromatin, which causes continuous repression of these genes. The active genes require intricate regulation and can be repressed when necessary by silencers, among others.

Insulators

The third important distal regulatory elements are insulators, also known as boundary elements. Insulators block the effect of other regulatory elements (enhancers and silencers) on certain genes. They are able to negate the effect from transcriptional regulation on nearby genes by blocking the

effect of enhancer-promoter communication. They can also prevent the spreading of repressive chromatin, by physically separating chromatin into independent structural domains.[24] For insulator activity the position is of vital importance, because they often form a physical boundary. They can however function in an orientation-independent manner. Due to the nature of their function, insulators are often found in dense genomic regions.[25] One protein often associated with insulators is CCCTC-binding factor (CTCF). CTCF can act as a repressor and activator, and is thought to prevent the spread of heterochromatin. The protein can prevent unwanted crosstalk between active and inactive genomic regions. In vertebrates, CTCF is the only known protein to mediate insulator function.[25,26] CTCF binding sites were found to be very conserved between different cell types, indicting functionality independent of cell type.[27]

Locus control regions

Next, there are the locus control regions (LCRs). These are groups of regulatory elements involved in the regulation of an entire locus or gene cluster. Each element has its own effect on the expression of specific genes and their collective activity is what defines LCR function, which is regulation of gene expression at the appropriate time and location. LCRs generally function like enhancers for these gene clusters. They can function from a distance, as all distal regulatory elements, and are orientation-independent. Because LCRs predominantly regulate entire gene clusters, they are crucial during certain stages of development, where gene clusters need to be regulated contiguously.[28]

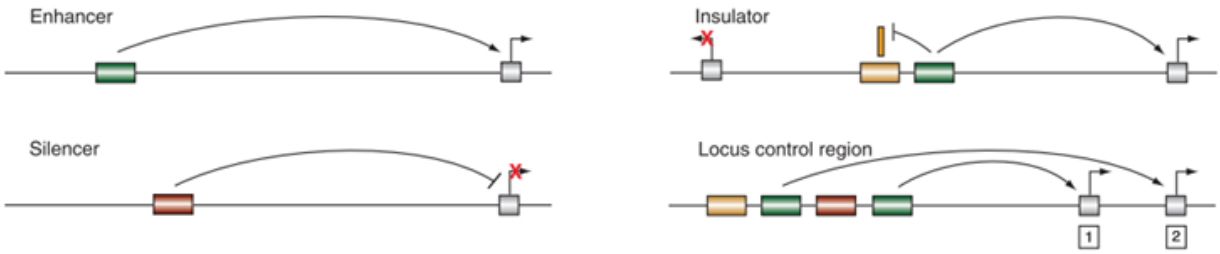

Figure 2. Activity of distal regulatory elements can influence gene expression.

Both enhancers and silencers regulate expression of their target gene from a distance, respectively increasing or decreasing expression. Insulators prevent this effect from happing on non-target genes by forming a blockade between the enhancer or silencer and these other genes. Locus control regions are groups of regulatory elements that together function to regulate a gene cluster. The combined regulation of certain genes is important for development and maturation of stem cells.

The combination of these distal regulatory elements, the transcription factors that bind to them, and the promoter elements, regulate the expression of genes. To discover the exact regulation of a certain gene, all regulatory elements and their associated transcription factors need to be identified. This identification has been going on for decades, but the process is tedious and new identification methods are required. Fortunately, the rise of the whole-genome era brought these new methods, one of which is the identification through histone modification patterns.

Histone modifications in regulatory elements have different biological functions

The transcription factors discussed above are often responsible for the deposition of modifications on the histones they attach to. The transcriptional machinery is also known to deposit modifications on the histones they pass.[29] This can happen either directly, or through the cofactors in the transcription factor complexes.[29] Many studies report that certain combinations of these histone modifications correlate with specific biological functions, such as transcription, replication or silencing. These observations led to the 'Histone Code' theory, which states that the combination of certain histone modifications specify a unique biological outcome to the region of the genome associated. Specialized binding proteins facilitate the specific function conferred by the histone modifications and give rise to downstream protein recruitment. Table 2 shows the different domains and the binding proteins for

several known histone modifications. Several protein domains that recognize specific histone modifications or histone modification patterns have been identified, such as Bromo, Chromo, Tudor or PHD.[30,31] It has been shown that changes in histone modification patterns are involved in the development and progression of certain human diseases, in particular cancer.[32]
Different histone modifications have different biological functions. Supplemental 1 provides a comprehensive list of all known and presumed functions of known histone modifications. Histone acetylation has been shown to be involved in gene activation.[3,11,33,34] Methylation of histones however, can indicate both gene activation or repression.[3,33,34,35] Well known among these, are the methylation of lysines 9 and 36 on histone 3 (H3K9me and H3K36me respectively). H3K36me can be found at histones in active genes and correlates positively with transcription levels. This modification is a remnant of the passing transcriptional machinery.[36] Methylation of H3K9 on the other hand, is present in silenced regions and is considered to be a hallmark for constitutive heterochromatin.[37] There are however studies that show that the silencing effect of H3K9me can be removed, while the modification is still present.[38] Potentially, the factors that enforce the silencing effect of H3K9me can be removed, while the modification is still present. It could also suggest that H3K9me needs multiple factors in order to confer its silencing function, which, when removed renders H3K9me incapable of its silencing capacity.[10,35]

Table 2. The Histone Code.[30,31]

Histone modifications are bound by specific transcriptional regulatory proteins by specific domains. Shown are certain histone modifications and the proteins that associate with them, grouped by the specific domains of the proteins that bind the modifications.

Histone Modification	Domain	Binding proteins
H3S10phos	14-3-3[46]	14-3-3 Family
H3S28phos		14-3-3 Family
H2A.X S139phos	BRCT[47]	53BP1, BRCA1, MDC1, NBS1
H3K9ac	Bromo[48]	BRD4, BAZ1B
H3K14ac		BRD4, BAZ1B, BRG1
H4K5ac		BRD4, BRDT
H4K12ac		BRD2, BRD4
H3K4me2/3	Chromo[49,50]	CHD1
H3K9me2/3		CDY, HP1, MPP8
H3K27me2.3		CDY, Pc
H3K36me		MRG15
H3K4me1	MBT[49]	L3MBTL1, L3MBTL2
H3K9me		L3MBTL1, MBTD1
H4K20me2		L3MBTL1
H3K4me3	PHD[51]	BPTF, ING, RAG2, BHC80, DNMT3L, PYGO1,
H3K9me3		JMJD2A
H4K20me		UHRF
		JMJD2A, PHF20
H3K4me3	Tudor[49,52]	TDRD7, JMJD2A
H3K9me		TDRD3
H3R17me		53BP1
H3K4me2	WD40[53]	WDR5
H3K9me		EED

Despite their diverse functions, it was shown that in some organisms not all histone modifications are essential to survival. For instance, deletion of the yeast methylation protein Set1 incapacitates the complex responsible for the trimethylation of H3K4, resulting in the depletion of H3K4me3. Despite this modification being a marker of active transcription, its depletion is well tolerated by yeast.[39] A similar study in *Drosophila* show comparable results; virtually normal regulation of transcription was found in absence of H3K4 methylation.[40] It is possible that these marks have a certain level of

redundancy with other active transcription markers. This would allow for survival after depletion of these histone modifications. One study on the necessity of histone modifications in yeast, showed that only 18 amino acids residues on all four core histones are essential for survival under laboratory conditions.[41] This might suggest that most histone modifications in yeast are not essential to survival. However, this needs to be thoroughly researched before any definite conclusions can be drawn. Though many of these residues seem unnecessary under laboratory conditions, this may be completely different under natural conditions. Interestingly, for mammals, an increasing amount of studies show the necessity of histone modifications.[42-44] It seems yeast requires less histones and histone modifications than mammals for survival.
Another example of life without histone modifications are the dinoflagellates. These are a diverse group of algae, and the only living eukaryotes known which lack histones all together. The DNA of dinoflagellates exists for a large part of non-protein coding DNA, which is thought to partly function in a structural manner, making up for the lack of histones. Dinoflagellates contain proteins similar to histones, which have been termed histone-like proteins. These however do not show any resemblance to the histone-like proteins present in prokaryotes. The function of these histone-like proteins in dinoflagellates has yet to be determined, but it is speculated that they are involved in transcription regulation. Unfortunately not much is known about the dinoflagellates, but they could provide great insight into the function of nucleosomes, histones and histone modifications, and the result of their absence.[45]

Transcription is regulated at many points, often through transcription factors. These TFs can bind regulatory elements in the genome to confer their enhancing or repressing activity on the expression of the gene. Deposition of histone modifications provide patterns that coincide with these regulatory elements, and provide attachment points for TFs containing specific histone modification pattern domains. Many patterns have been identified so far, mostly on promoters and enhancers. These patterns will be discussed in the next chapter.

Chapter 2 Histone modifications on known regulatory elements

The ENCODE project

Analysis of histone modifications in the human genome has revealed that specific combinations of these modifications form patterns that can identify regulatory elements.[33] Experiments to specify histone modifications per regulatory element have been done in several organisms, among which are *Saccharomyces ceravisiae* and *Drosophila*, and mammals such as mouse and human. These experiments were able to identify specific patterns of histone modifications for several regulatory elements in all these model organisms.[35,54,55] Comparative genomics have shown that these patterns are at least partially conserved from yeast to humans (table 3). Between mouse and human, the level of conservation of histone modifications patterns is quite high. DNA sequences of these sites however do not show the same extent of conservation. This indicates that relying solely on comparative genomics may be insufficient to identify the complete extent of regulatory elements.[56]

Mapping of histone modification patterns in the human genome will help to identify the regulatory elements that escaped discovery through comparative genomics. In order to collect and supply information regarding the identification of functional elements in the human genome, an international project has been set up. This project, the Encyclopedia of DNA Elements (ENCODE) project, has examined approximately 1% of the human genome for functional elements using a multitude of existing techniques.[57]

Table 3. Conservation of histone modifications in different model organisms.

Histone modifications are generally well conserved throughout evolution and some can be found from *Drosophila* to humans. Yeast however appears to lack most histone modifications. The presence or absence of several modifications on histone H3 are shown for *Saccharomyces ceravisiae* (yeast), *Drosophila*, mouse and human.

		Yeast	*Drosophila*	Mouse	Human
H3 Lysine 4	Acetyl	-		+/-	+
	Monomethyl	+++	+++	++	++
	Dimethyl	++	++	+	+
	Trimethyl	++	+	+/-	+/-
H3 Lysine 9	Acetyl	+		+/-	+/-
	Monomethyl	-	++	++	++
	Dimethyl	-	+	++	++
	Trimethyl	-	++	+++	+++
H3 Lysine 27	Acetyl	+++		+	+/-
	Monomethyl	-	+	++	+++
	Dimethyl	-	+++	+++	+++
	Trimethyl	-	+++	++	++
H3 Lysine 36	Acetyl	+/-		-	+/-
	Monomethyl	++	++	++	++
	Dimethyl	++	++	++	++
	Trimethyl	++	+/-	+/-	+/-
H3 Lysine 56	Acetyl	+++		-	+/-
	Monomethyl	-		-	+
	Dimethyl	-		-	-
	Trimethyl	-		-	+++
H3 Lysine 79	Acetyl	-		-	+/-
	Monomethyl	++		+	+
	Dimethyl	++	++	+	+
	Trimethyl	++	-	-	-

- *absence of histone modification*

+/- up to +++ *degree of conservation of global levels between different organisms*

Techniques to identify regulatory elements

ChIP

The most used technique to identify histone modifications is Chromatin Immunoprecipitation (ChIP). ChIP experiments require specific antibodies against modified histones. These antibodies facilitate extraction of the modified histones together with their associated DNA. Specificity of the antibodies in these experiments is crucial. One study investigated the specificity of several antibodies against histone modifications, and found that approximately 20% was not effective.[58] This displays the importance of verifying the effectiveness of antibodies used, and suggests that experiments performed with these ineffective antibodies could have flawed results.

ChIP experiments are generally followed by either a microarray (ChIP-chip),[26] sequencing (ChIP-seq)[54] or a serial analysis of gene expression (ChIP-SAGE).[54] These three different techniques all have their advantages and disadvantages. The resolution of the results differ between the three follow-up experiments. When micrococcal nuclease is used for ChIP-seq, it is possible to analyze single nucleosomes, providing the highest resolution. For ChIP-chip the resolution depends on the size of the chromatin fragments after digestion, and on the probes present on the microarray. The resolution for ChIP-SAGE depends on the frequency of the restriction sites in the DNA fragments. ChIP-seq is considered the most cost-effective for whole genome sequencing. In some cases ChIP-seq is combined with tiling oligonucleotide assays to identify regulatory regions in previously unexplored DNA.[56]

FAIRE

Another frequently used technique is Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE). In FAIRE experiments the chromatin is cross-linked by formaldehyde and sheared, similar to ChIP experiments. FAIRE was found to specifically target DNA coincident with TSSs, active promoters and DNase I hypersensitive sites. This coincides with the fact that formaldehyde cross-linking is more efficient in nucleosome-bound genomic regions. Nucleosome-depleted regions are often poorly cross-linked. After treatment with formaldehyde, the non-cross-linked DNA is segregated through phenol-chloroform extraction. This sequesters protein-associated DNA in the organic phase and leaves unbound DNA free in the aqueous phase. The DNA present in the aqueous phase is subsequently sequenced. This segregation results in the seeming preference of FAIRE for active genomic regions, which are often linked with nucleosome-depleted regions.[59]

DNase I hypersensitivity

Probably the most useful feature of all regulatory elements in many organisms, is the presence of DNase I hypersensitive sites (DHS). These sites possess heightened sensitivity for cleavage by the enzyme DNase I, compared to other genomic regions. This heightened sensitivity is derived from the fact that these sites are mostly nucleosome-depleted, allowing for easy access and cleavage by DNase I. DHSs are precisely localized to regulatory elements, and are therefore extensively used to identify and demarcate any regulatory element in the genome.[60] Digestion with DNase I followed by sequencing will provide the DNase I hypersensitive regions. Genome-wide mapping of these sites has been performed in multiple organisms to identify regions involved in transcriptional regulation. However, these sites do not show the nature of this element, which will have to be identified through different techniques. DNase I hypersensitive sites are flanked by histones, which can acquire the modifications specific for this region. This can allow identification of the region of the regulatory element, and its nature.[60]

Further data analysis

In order to verify the results obtained from ChIP or FAIRE experiments, several approaches are being used. Results are compared with existing genome maps to verify previously discovered regulatory elements as a control. Further verification can be achieved by comparing with results from the ENCODE database, for instance by using a multivariate Hidden Markov model.[10] Using the newly discovered

regulatory element in reporter assays, or introducing point mutations decreasing its activity, are other methods used to verify the function of the newly discovered elements.

Gene ontology
One approach occasionally used to analyze the effects of the histone code on downstream factors is gene ontology. This techniques attempts to categorize genes or gene products in different functional groups, by using several properties (such as location, temporal expression, or molecular function). In the case of histone modifications, this can be applied to annotate functional properties to patterns of histone modifications. Gene ontology will attempt to categorize the genes associated with certain patterns into groups representing different biological functions, such as replication.[61]
Gene ontology used on patterns of histone acetylation in yeast showed that the combination of H3K18ac, H3K9ac and H3K27ac was linked to protein synthesis genes.[54] When used on human T-cells, this approach has indicated the enrichment of 17 different histone modifications in active genes involved in cellular physiology and metabolism.[3]

All these techniques have been used extensively, resulting in many known histone modification patterns for coding regions, promoters and enhancers.[62] In lesser extent research was done on insulators, silencers and locus control regions.[59] Large-scale research on methylation and acetylation of histones in eukaryotes have shown that these findings are generally consistent between different model organisms.[35,56] Results are often compared with the data from the ENCODE project, which contains information about the histone modification patterns on regulatory elements in approximately 1% of the human DNA.[57] This comparison often results in confirmation of the newly discovered histone modification patterns or regulatory elements.

Active coding regions in mammals are generally defined hypoacetylation and methylation of H3K36
Much has been discovered about the histone modifications present in active coding regions. Best known is the acetylation of lysines. This can occur on all histones and was shown to be associated with activation of transcription. In yeast, active coding regions show enrichment for acetylation on many locations, such as H2aK7, H3K9, H3K14, H3K18, H4K5 and H4K12, while inactive regions are often associated with hypoacetylation. The acetylated lysines tend to be localized on histones towards the 5’ end of transcriptionally active genes.[3] Acetylation in yeast is important throughout its entire genome due to its transcription-enhancing activity. It is thought that nucleosomes act repressive on transcription initiation in yeast. Acetylation of nucleosomes is proposed to relieve the repressing effect of nucleosomes on transcription. This would also explain the preference of acetylation to the 5’end of active genes.[63]
A large-scale study for human histone acetylations showed acetylation on all histones correlated with gene activation and expression.[64] Activation of transcription via acetylation of the DNA is generally thought to occur through neutralization of the positive charge of the lysine residues, reducing their affinity to DNA, and opening up the chromatin structure.[3,64] Promoters of active genes are thus highly acetylated. Coding regions however, show hypoacetylation to prevent off-target transcription initiation. This is different from acetylation in the yeast genome, although in yeast acetylation also localizes more to the 5’ end of coding regions. Interestingly, hyperacetylation was shown to be involved in retaining cells in an undifferentiated and pluripotent state.[65]

Methylation of histones is more complex and can have both a positive and negative influence on transcription, dependent on the position and state of the modification and the cell. There are twenty-four known methylation sites on histones, all of which can contain up to three methyl groups. Coding regions in yeast have been shown to contain H3K36 trimethylation and H3K4 methylation. The latter was shown to have an increasing degree of methylation (mono-, di- and tri-) in a 3’ to 5’ direction. Trimethylation of H3K36 is caused by the methyltransferase Set2, a protein known to be associated with elongation of RNA Polymerase II. In line with this, was the discovery that the level of H3K36me3

correlates with the level of transcription. High expression of a gene increases H3K36me3 along that gene body, indicating that transcription is involved in the establishment of this histone modification.[35] The histone deacetylase complex Rpd3S recognizes H3K36 methylation and specifically deacetylates these coding regions, to create the deacetylated coding regions described above.[66] This indicates that methylation of H3K36 could be vital for the organism, because it inhibits hyperacetylation of coding regions and therefore prevents unwanted transcription initiation. Methyltransferase Set1 is recruited to 5' regions of genes actively transcribed by RNA Polymerase II, and this protein is responsible for the methylation of H3K4.[67]
In conclusion, active coding regions in yeast are thus associated with hyperacetylation, and methylation of H3K36 and H3k4. In mammals these regions are marked by hypoacetylated gene bodies, H3K36me and H3K4me.

Promoters are defined by nucleosome depletion and H3K4 methylation
Just like coding regions, promoters have been extensively researched for matching histone modification patterns. In yeast and *Drosophila* it was shown that many active promoters were partially depleted of nucleosomes. This increases accessibility of the DNA in this region, making it more accessible to the transcriptional machinery.[36] This depletion of nucleosomes at promoter regions is considered to be true for mammals also. Research in HeLa cells has shown that the core histone H3 seems depleted from actively transcribed promoters, but this remains to be thoroughly researched.[68] For nucleosomes that are present in promoter regions, it was found they often contain histone H2A.Z instead of the standard histone H2A.[68] One study in *Arabidopsis thaliana* indicated a direct role for histone variant H2A.Z in averting DNA methylation at these sites. They show general DNA methylation at sites of H2A.Z depletion.[69] DNA methylation is often considered to coincide with heterochromatin or silenced genomic regions. Averting DNA methylation at active promoters is therefore essential.

Table 4. Promoter histone modification patterns in different model organisms.
Histone modification patterns that have been identified in promoters in humans, *Drosophila* and yeast, grouped in patterns found in promoters yielding high, intermediate and low expression.[44]

Gene expression	Human CD4+ T-cells	Human ES cells	*Drosophila*	Yeast	Yeast (single-nucleotide mapping)
High	H2A.Z backbone*, H2aK9ac, H2bK5me1, H3K79me1/2/3, H4K12ac, H4K16ac, H4K20me1	H3K9ac, H3K14ac and H3K4me3	Hyperacetylated for H3 and H4; H3K4me and H3K79me	H3K18ac, H3K9ac, H3K27ac	H3K18ac, H4K12ac, H3K9ac, H3K14ac, H4K5ac, H3K4me3
Intermediate	H4K16ac, H3K36me3				
Low	H2A.Z, H3K4me1/2/3, H3K9me1, H3K27me3	H3K4me3 (low levels)	Hypomethylated and deacetylated residues	H3K8ac, H4K16ac, H2bK11ac, H2bK16ac (H3k23ac, H2aK7ac)	Hypomethylated and deacetylated residues

* H2A.Z backbone = H2bK5ac, H2bK12ac, H2bK20ac, H2bK120ac, H3K4ac, H3K4me1/2/3, H3K9ac, H3K9me1, H3K18ac, H3K27ac, H3K36ac, H4K5ac, H4K8ac and H4K91ac on histone H2A.Z

Most histones in promoters contain a specific pattern of methylation and acetylation. At the TSS they show enrichment for H3K4 dimethylation and trimethylation, and acetylation of histone H3 (mainly H3K9, H3K14 and H3K18). Other specific acetylations found at the TSS are located at H2aK9, H2bK5, H3K27, H3K36 and H4K91. Downstream from the TSS, promoters show a slight enrichment of H3K4 monomethylation and acetylation of H2bK12, H2bK20, H2bK120, H3K4, H4K5, H4K8, H4K12 and H4K16. Table 4 shows a comparison of known histone modification patterns in different model organisms.[56,62,64,65] Just like in yeast, acetylation in mammals also creates a more open chromatin formation, explaining the high level of acetylated histones near active promoters. Methylation of H3K4

is often considered to be associated with RNA Polymerase II presence. Both are thought to be linked to open chromatin formation, possibly by the interaction with chromatin remodeling factors.[70]
Obviously, the histone modification pattern for promoters at non-active genes is somewhat different. At these promoters a slight enrichment of both di- and trimethylation at H3K4 was observed on histones centered in the TSS.[10] The best characteristic of non-active gene promoters however, is the absence of acetylations at most histones. In some repressed gene promoters in yeast, there does seem to be some level of acetylation at H4K8 and H4K16, but this is negligible compared to the hyperacetylated promoters of active genes.[62]
In general, active promoters are easiest recognized by hyperacetylation, combined with H3K4me2/3, whereas inactive promoters show hypoacetylation and slight enrichment of H3K4me1/2.

Enhancers show different histone modification patterns for active, inactive and poised states
From al distal regulatory elements, enhancers have been studied most extensively. The histone modifications observed on enhancing regulatory elements was shown to vary between active, inactive and poised enhancers. The best known histone modification at enhancers is the monomethylation of lysine 4 at histone H3 (H3K4), along with the depletion of trimethylation at H3K4. One study showed that over 20% of enhancers contain the core histone H2A.Z instead of H2A, and have an enrichment for monomethylation of H3K4, monomethylation of H3K9 and acetylation of H3K18.[3,35,71] Both active and poised enhancers were shown to contain H3K4me1, which is known as an active transcription mark. On poised enhancers this facilitates fast activation when required. Enrichment of H3K27 acetylation can be found on active and engaged enhancers and is known to demarcate them. Interestingly, this modification is not found on poised enhancers, providing a good mark for proper distinction between poised and active enhancers.[71,72] Table 5 shows a summary of histone modifications found in active, poised and inactive enhancers in different organisms. Generally these modifications are the same between human, mouse, *Drosophila* and yeast.

Table 5. Enhancer histone modification patterns in different model organisms.
Histone modification patterns that have been identified in enhancers in humans, *Drosophila* and yeast, grouped in patterns found in active, poised and inactive enhancers.

Gene expression	Human cells[74,75]	Mouse[54]	*Drosophila*[55]	Yeast[76]
Active	H2A.Z, H3K4me1/2, H3K9me1, H3K18ac, H3K27ac, H3K36me1, H4K20me1 (H3K4me3 depletion)	H3K4me1, H3K27ac	H3K4me1, H3K27ac	H3K4me1, H3K27ac
Poised	H3K4me1/2, H2K9me3, H3K27me3	H3K27me3	H3K27me3	
Inactive	(H3K4me1 depletion), H3K27me3		H3K27me3	

A large-scale study to the methylation patterns on different regulatory elements has shown that enhancers often include combinations of the modifications H3K4me1, H3K4me2, H3K9me1, H3K36me1 and H4K20me1.[71] They show that not all enhancers contain the same pattern, but most contain a combination of these PTMs. To increase our knowledge of enhancer localization and activity, it is necessary to investigate the patterns found on enhancers and their connection to spatial and temporal expression.[71] Important in the research to enhancer activity is the fact that enhancers work cell type specific.[72] Specific enhancer histone modifications, such as H3K4me1, are often only present on enhancers that are active in that cell. This shows that enhancers are important for cell type specific expression, regulating transcription of cell type specific genes.[72]
Conclusively, the distinction between active, poised and inactive enhancers can be made by paying attention to only a few histone modifications (methylation of H3K4; and acetylation of methylation of

H3K27), which is important because enhancers act cell-type specific and need to be mapped for each cell type separately.

Histone modifications for other regulatory elements are less well defined

Compared to promoters and enhancers, relatively little is known about the histone modifications associated with insulators. Insulators are mostly identified using the transcriptional repressor CTCF, which is involved in regulation of chromatin structure, and associates with insulators. So far, CTCF is the only known protein that binds to insulator elements in mammals.[23] In *Drosophila*, besides their homologue dCTCF, several other proteins have been shown to bind insulators, such as Su(Hw), BEAF-32 and GAGA Associated Factor (GAF). These proteins all bind to regions of insulators that appear to have a reduced nucleosome density. Low nucleosome density is caused by loosely packed chromatin, which corresponds with regions of active transcription. These proteins therefore seem to be actively contributing to the enhancer-blocking and chromatin-changing functions of insulators near active genes.[26]

Because ChIP experiments using CTCF as bait are very reliable to identify insulators, not much attention has been given to specific histone modifications at these regulatory elements. It has been proposed that ubiquitination of histones may indicate the presence of insulators in the genome.[77] One study shows that proteins associated with HS4, the insulator element located as a boundary between the β-globin gene cluster and a condensed chromatin domain, recruits enzymes that mediate chromatin-changing histone modifications.[77] Some of these enzymes are known to ubiquitinate histone H2B, and it was indeed found that some nucleosomes at this insulator are ubiquitinated. It is believed that this is done by the enzyme RNF20, because depletion of this enzyme causes compromised boundaries and spreading of heterochromatin to active gene regions.[77] Besides ubiquitination, it is also speculated that boundary elements are enriched for H3K9me1 and devoid of acetylations.[35] These findings however are still very preliminary and need to be further investigated.

Similar to insulators, silencers are also underrepresented in the search for histone modification patterns, but this will be further detailed in chapter 3. For locus control regions there are no specific histone modification patterns known, because they consist of several regulatory elements that together regulate the transcription of a gene cluster. These regulatory elements are all defined by their own specific histone modification patterns. Because they resemble enhancers, mostly found in locus control regions are histone modification patterns found on enhancers. Generally little is known about histone modifications on distal regulatory elements other than enhancers. This is either because there are other reliable methods for identifying these elements, because they don't have specific histone modification patterns, or because they have not been studied extensively enough.

Bivalent domains contain both active and inactive histone modification

An interesting observation made from large-scale genomic experiments in the human genome, is the discovery of bivalent domains.[10] These are particular sequences in the genome, mainly found in developmental genes in embryonic stem cells, where nucleosomes contain both the repressive H3K27 trimethylation and the activating H3K4 trimethylation. These genes tend to have a very low or no expression, indicating that the repressive modification is dominant. The trimethylation of H3K4 keeps the gene poised for transcription for when the repressive mark is removed. This is likely to happen during development. The bivalent imprint disappears, and either the repressing or activating histone modification is left behind. Genes necessary for the developing or developed cell become active, whereas all other genes lose their H3K4 trimethylation and become fully repressed. Some genes were found to keep their bivalent status, although it has not yet been shown why this happens, or if there is any connection between the genes that retain both histone modifications.[78]

Genomic mapping can increase our knowledge regarding transcriptional regulation in different organisms

Genomic mapping experiments are increasingly being performed in an effort to create a complete overview of cell regulation at a transcriptional level.[54,55] Identification of regulatory elements through histone modifications can help reconstruct regulatory networks. Some transcription factors use specific histone modifications to bind to, or to guide them to a binding site.[14,15] Combining histone modification patterns with transcription factor binding sites can provide more insight in the regulation of transcription. Some transcription factors have been observed to bind outside dictated promoter regions.[79] When overlaying histone modification pattern maps with transcription factor binding site maps, these particular regions can be identified. This could provide new insight into non-promoter binding sites, identify new promoter region characteristics, or help assign new functions to transcription factors or these non-promoter binding sites. Several of these regions, which were thought to be enhancers, have thus far been identified as distal promoter regions.[79] This offers a different view on the localization and possibly function of promoters, and needs to be investigated thoroughly.

Known histone modification patterns are being used to increase our understanding of function and positioning of regulatory elements in the human genome. Identification of new regulatory elements is facilitated by these patterns. Established histone modification patterns can also aid in the identification of alternated patterns in cells in diseased state. Many variants of cancer are known to be influenced by epigenetic modifications, which can be identified by comparing histone modification patterns to those of healthy cells.[80]

Chapter 3 Histone modifications on silencers and silenced DNA

The global efforts of the ENCODE project have greatly assisted in the annotation of functional elements to non-coding regions in the human genome. Particularly for positive regulatory elements, such as enhancers and promoters, insight has greatly increased with regard to their position and function over the past years. Much has been discovered on specific characteristics that enable easy identification of new positive regulatory elements, such as specific histone modification patterns. Despite the great increase in our knowledge on these elements, the research on negative regulatory elements, such as insulators and silencers, has greatly lagged behind. These elements are bound to be equally abundant in the genome, but only a small amount has been discovered or extensively described.[26,59] Particularly silencers are underrepresented in literature.

There are several different states of silenced chromatin. Heterochromatin consists of generally inactive DNA, tightly packaged into the 30-nm fiber, as described in chapter 1. Heterochromatin can be subdivided into constitutive heterochromatin and facultative heterochromatin. Constitutive heterochromatin is known to form structural functions, such as telomeres and centromeres, and often consists of repetitive DNA sequences. These sequences are generally packed into heterochromatin in every cell in an organism. Two well-known examples of constitutive heterochromatin are centric heterochromatin and pericentric heterochromatin. Centric heterochromatin is present on the centromeres of chromosomes, directly flanked by pericentric heterochromatin. They are thought to play a role in chromosome segregation. It has also been proposed that pericentric heterochromatin is involved in repression of genes.[81]
Facultative heterochromatin consists of deliberately silenced genomic regions. These regions are packed into heterochromatin to prevent the expression of the genes present. This will result in different sequences being packed into facultative heterochromatin between different cell types.[82]
Finally, euchromatin can also be actively silenced by silencing regulatory elements, preventing the transcription of genes.

Silenced regions are indicated by methylation of H3K9 and hypoacetylation
Despite the gap in knowledge on histone modification patterns in silencing elements, relatively much is known about histone modifications in silenced genomic regions. Transcriptional gene silencing is the result of condensed heterochromatin, disabling the transcriptional machinery to access the DNA. Generally associated with transcriptional silencing is DNA methylation, the histone modifications H3K9me2, H3K9me3, H4K20me, and histone deacetylation. Aside from these well-known marks, there are also several other modifications associated with silenced genomic regions.[3,56,67]

Histone modification associated with silenced genomic regions can vary between species. Methylation of H3K27 for example is not found in yeast, but can be present at repressed genes in *Drosophila* or mammals. And while methylation of H3K9 has been shown to be essential for regulation and maintenance of DNA methylation in several species, among which the plant model *Arabidopsis thaliana*,[83] yeast does not possess this specific modification.[84]
In fact, yeast does not contain any of the marks that have specifically been associated with transcriptional repression. DNA methylation was shown to be present in yeast in very low percentages compared to mammals.[85] Silenced regions may however be distinguished by observing the absence of positive histone modifications. Absence of these histone marks could provide high-affinity binding sites for Silent Information Regulation (SIR) proteins, which are histone deacetylases recruited to the DNA to invoke genomic silencing. The SIR complex causes tightening of the chromatin and thus silencing through deacetylation of the histones.[86] Generally this suggest that for yeast, acetylations are the most important histone modifications, regulating transcription activity.
Repressive chromatin marks in *Drosophila* include methylation of H3K9, H3K27 and H4K20. Heterochromatin can be defined by all methylation states of these three lysines. Pericentric

heterochromatin for instance, is indicated by the presence of H3K9me1, H3K9me3, H4K20me3 and all methylation states of H3K27. Though H3K9me was shown to be present on silenced genomic regions in *Drosophila*, it is not as prominent as in mammals. Another interesting observation in *Drosophila* is that methylation of H3K27 can be present at both heterochromatin and silenced euchromatin, which has thus far not been observed in other organisms. Both cases involve silenced genomic regions, but in organisms other than *Drosophila*, these regions contain distinct marks.[36,37,55] One could compare this with the bivalent domains in mammals, which also contain methylation of H3K27. Bivalent domains however are not defined as heterochromatin, but as poised chromatin, which can become either heterochromatin or euchromatin upon the right interactions.[10,78]

In mammals, constitutive heterochromatin is characterized by the histone modifications H3K9me3, H3K27me1 and H4K20me3, general histone deacetylation and DNA methylation.[87] One study in HeLa cells shows that ubiquitination of H2aK119 is also involved in gene silencing. This needs to be further verified, but appears to be involved in silencing through the Polycomb Group proteins.[88] Deubiquitination of H2B was also shown to be associated with gene silencing.[89] Establishment of silenced genomic regions by the formation of heterochromatin was shown to follow several specific steps. First histones are deacetylated by HDACs, which often co-occurs with the methylation of H3K9. Next Heterochromatin Protein 1 (HP1), which specifically targets methylated H3K9, binds to the histones through its chromo-domain. HP1 causes chromatin transformation, resulting in formation of heterochromatin. Interestingly, it was found that HP1 can be removed from H3K9m3 sites upon phosphorylation of H3S10. This indicates that the silencing effect can be removed, while the histone modification is still present,[90] resulting in active genes with H3K9me3.[91]
In conclusion, the best known histone modifications for heterochromatin are methylation of H3K9 and H3K27, often combined with some other PTMs which can vary between different organisms. H3K9 attracts the protein HP1, which subsequently causes transformation of chromatin to heterochromatin.

Interaction between silencing histone modifications and DNA methylation
The best known marker for silenced genomic regions is the methylation of the DNA. DNA methylation is the methylation of cytosines, causing increased difficulty for TFs to interact with these stretches of DNA. Recent investigations have directed attention to interactions between DNA methylation and silencing histone modifications. DNA methyltransferases can methylate cytosines within the promoter of genes, which will subsequently bind proteins like MeCP2. These proteins recruit nucleosome modifying protein complexes to silence the target region. In some cases these modifications are able to recruit proteins that establish the same modifications, creating a certain level of redundancy and enabling the cell to maintain this silenced state throughout several cell divisions. One study by Rothbart *et al.* shows a mechanistic link between silencing histone modifications and DNA methylation, which is important for the inheritance of DNA methylation patterns during cell division.[92] They show that the DNA methyltransferase DNMT1, which is known to be important for the inheritance of DNA methylation, is recruited to methylated CpG-dinucleotides by the E3 ubiquitin ligase UHRF1. UHRF1 was shown to preferentially bind di- and trimethylated H3K9 even in the presence of H3S10 phosphorylation, a known mitosis marker able to prevent binding of proteins in its surroundings. This indicates that UHRF1 can overcome this aversion of H3S10 phosphorylation during mitosis in order to regulate proper DNA methylation inheritance.[92] Another study shows that at pericentric chromatin, several histone methyltransferases, such as Suv39h, are required for H3K9 methylation and DNA methylation.[93] Generally, besides the physical impediment of DNA methylation for transcription factor binding, DNA methylation may also interact with silencing histone modifications to form heterochromatin.

Silencers have been described in different model organisms
Besides the formation of heterochromatin, another mechanism of silencing is required in cells. Silencers are involved in the silencing of genes in euchromatin. Unfortunately, little is known about

silencers in mammals. Silencers and their function have been studied to a greater extent in non-vertebrae such as *Drosophila* and yeast. One study towards the element required for CD4 silencing in mouse T cells, compared the discovered silencer with its homologue in *Drosophila* and yeast.[94] Comparing the mouse CD4 silencer with yeast, they indicated that there are several sites acting with redundant function just as the HML and HMR mating type locus silencers. And just like in *Drosophila*, the mouse CD4 silencing activity can be partially enhanced by increased levels of HP-1β expression.[94] But while in yeast and *Drosophila* more silencing elements have been identified and their function has been investigated in more detail, these investigations do not shed more light on the presence of specific histone modification patterns in silencers.

Silencers are often identified using reporter assays, containing a strongly expressed gene (either with or without its connected enhancer) as a reporter for functional silencing. When the researched genomic region is able to effectively diminish the transcription of the reporter gene, it is likely to contain a silencing element. Verification of potential silencing elements can be done in several ways, using the techniques discussed earlier. Introduction of a frame-shift mutation in the suspected region can disable the silencing function. A point mutation to disable transcription factor (repressor) binding can accomplish the same goal. If generally repressed genes are no longer repressed after introduction of the mutation, it is likely that the suspected region indeed contains silencing properties. Comparing the suspected region with other datasets, such as those gathered in the ENCODE project, can also provide more information on its fuction.[95] Like all regulatory elements, silencers contain DNase I hypersensitive sites, which can aid in the discovery of new silencing elements.[3] Despite the fact that various silencer identification methods have been developed and applied, this research is not widespread and not many comparative investigations have been performed in order to find a general consensus between most silencer regions. This gap in research could explain the lack of knowledge of common histone modifications in silencing regulatory elements, provided these patterns exist.

Silencer-specific histone modifications are sparsely described in literature

Fortunately, there has been some research on the presence of specific histone modifications in silencers. One study by H. Petrykowska and coworkers shows the identification and characterization of silencers and insulators in the greater *CFTR* locus.[59] They used a luciferase reporter assay to determine the functionality of 47 target regions from the greater *CFTR* locus. Most of these regions were located in introns, and approximately two thirds were shown to include conserved noncoding regions, suspected to be regulatory elements. A region was determined to contain silencing abilities when the luciferase activity in presence of this region was lower than the luciferase activity of the control, despite the presence of an enhancer.[59]

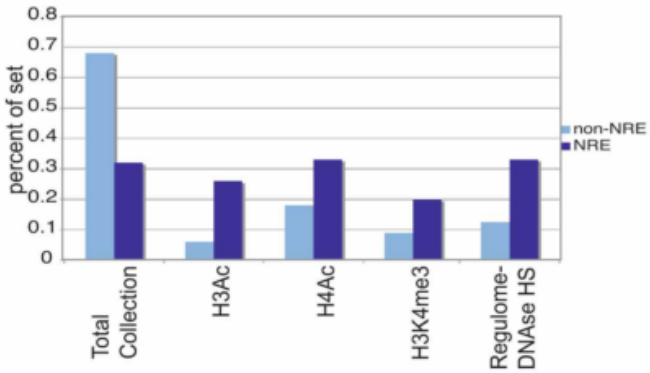

Figure 3. Comparing non-NRE and NRE to find specific histone modifications

When comparing histone modifications on all negative regulatory elements (NRE) with those of non-NRE, several histone modifications were found to be enriched for NREs. Acetylation of histones H3 and H4 and trimethylation of H3K4 seemed enriched for NREs, just like DNase I hypersensitive sites. Unfortunately the groups were too small to perform proper statistical analysis and indicate significance.[59]

From these 47 selected regions, ten proved to contain strong silencing abilities (and six contained insulator capabilities), most of which acted in an orientation-dependent manner. From these ten regions, 70% was found to be capable of silencing in more than one cell line. These negative regulatory elements (NREs), both the silencers and the insulators, were grouped and compared with the other regions. Using ENCODE data distinct differences between both groups were searched. They showed that the NREs are enriched for histone acetylation at histone H3 and H4, compared with the other group (Fig. 3). They also found a slight enrichment of trimethylation of H3K4 and DNase I hypersensitive sites in the NREs. They did not find any other histone modification differences between both groups using the ENCODE database. Unfortunately, they were unable to accurately quantify and state the statistical significance of this comparison, due to the small sizes of both groups.[59]

Studies in yeast have revealed that a defined set of proteins is required for epigenetic silencing. These proteins are known to interact with parts of the tail of core histones H3 (amino acid 3-20) and H4 (amino acid 16-29). One of these residues, H4K16, is an important target for acetylation in actively transcribed genes. Lack of acetylation on this site allows these proteins to interact and induce silencing in this region. This seems contradictive with the results discussed previously, where hyperacetylation was associated with some silencers. Both possibilities however are based on presumptions from a single research and need to be investigated further before any conclusions can be drawn.[84] Aside from this, both investigations were performed in different organisms (yeast and humans) and, as discussed earlier, yeast has a distinct acetylation pattern compared with other model organisms.

Research on silencers and silenced genomic regions will have to increase in order to improve our understanding of how gene expression is repressed. Considering this repression is the default state of most genes in eukaryotic cells, it is particularly odd that we still know so little about the elements that cause this repression. Identifying histone modification patterns will be of use in the discovery of new silencing regions, and can benefit research on the method silencers use for the repression of specific genes. Finding interactions between different silencing elements, as with histone modifications and DNA methylation for silenced regions, can provide insight into transcription regulation in general. These new insights can subsequently be used to identify tumor suppressors, or medicines in other diseases affected by histone modifications, such as personality disorders.

Chapter 4 Applications of epigenomic patterns

The identification of histone modification patterns and other epigenomic characteristics of transcriptional regulation have given us new perspectives to identify novel regulatory elements, and implement these in our current models. These patterns have been a tool for identification of regulatory elements in the genome of model organisms in many studies.

Epigenomic patterns have been used to identify novel regulatory elements in different organisms

Identification of DNase I hypersensitive sites (DHSs) has led to the discovery of many novel and previously identified regulatory sites in the human genome. All discovered DHSs in one study were subjected to extensive investigation to identify them as regulatory elements.[60] To annotate promoters they compared the DHS map with the results from ChIP-seq experiments for H3K4me3, resulting in the identification of over 44,000 novel promoter regions in 56 different cell types.[60] In total 99.5% of these proved to be genuine promoters when subjected to identification using different promoter characteristics. For annotation of distal regulatory elements, DHSs were compared with DHS patterns found in the well described β-globin locus control region. This revealed many enhancer regions, some of which were used for verification through reporter assays.[60]

Another study used epigenomic patterns to annotate genomic regions in the mouse genome. Through ChIP-seq for RNA Polymerase II, the protein CTCF, and the histone modifications H3K4me3, H3K4me1 and K3K27ac, they identified regulatory elements in the mouse genome in different adult and embryonic cell lines. They were able to identify over 295,000 non-redundant regulatory elements, including over 53,000 promoters (enrichment of H3K4me3 and DNA Polymerase II binding sites), over 234,000 enhancers (enriched for H3K4me1 and H3K27ac), and over 111,000 insulators (CTCF binding sites). These potential regulatory regions were used for validation experiments. Promoters were compared with known promoters, validating many sites and identifying over 13,000 potential novel promoters. Reporter assays on some, validated that approximately 85% showed significant promoter activity. The enhancer and insulator regions were also subjected to validation assays, verifying many previously annotated regions and discovering many novel ones.[54]

Similar experiments are being conducted on many different stages of development, in different cell types, and in many organisms. The annotation of regulatory elements continues to be a topic of great interest and importance. Though many promoters, enhancers and insulators are being detected, the discovery of silencers is lagging behind greatly. To fully understand transcriptional regulation, these regions will have to be accounted for too. Unfortunately, there are no useful epigenomic patterns known yet to easily identify novel silencers.

Epigenomic patterns have been used to identify differences in cancer cells

Despite this gap in our knowledge on silencers, patterns for other elements are being used extensively to discover differences between cells in a healthy and diseased state, often focusing on cancer development. One study has used ChIP-seq (for H3K4me1, H3K4me3, H3K9ac and H3K14ac) to identify regulatory elements in human breast cancer cells, and compare these with healthy breast cancer cells.[72] This led to the identification of divergent histone modifications on many differentially expressed genes in the breast cancer cells. Disturbed epigenetics is considered to be an early event in the development of cancer, and often cause of genomic instability, one of the hallmarks of cancer. It is considered that epigenomic characteristics are heritable throughout several cell divisions, contributing to the development of cancer.[96]

Epigenomic changes have been reported for many different forms of cancer. In acute leukemia and medulloblastomas, oncogenes were shown to be enriched for H3K9ac and H3K4me2.[97] In non-small-cell lung cancer levels of histone modifications were used to predict survival. Global enrichment of H3K4me2, H2aK5ac and H3K9ac was shown to correlate with higher survival rates.[98] In one study, Seligson and coworkers show the importance of investigating histone modification patterns.[99] They used histone modifications for the prognosis of prostate cancer. When studying the modifications

separately, no significant results were obtained. However, when analyzing specific modifications patterns, they were able to find a prognostic tendency. They observed that the levels of histone modifications between benign, pre-neoplastic and cancerous cells differ enough for distinction between these states.[99] Cancerous cells showed global enrichment of several histone modifications (H3K18ac, H4K12ac, H4R3me2 and H3K4me2). Distinction between benign and pre-neoplastic could be done using H3K4me2 and H3K18ac levels. High levels of H3K4me2 occurred in benign cells, whereas low levels of both H3K4me2 and H3K18ac indicated pre-neoplastic cells with higher chance of cancer reoccurrence.[99]
The ability to distinguish between different states of cancer development based on histone modifications may provide new possibilities for cancer diagnostics. It may increase ease of early diagnosis of the developmental state of the cancer. This could be used for application of the right treatment at the proper time, and can be vital for the survival of a patient. It is possible that the discovered patterns of histone modifications are also applicable to other tumor types, increasing diagnostic possibilities in many types of cancer.

Several histone modifications have been linked to cancer development. In many cancer cells, reduction of the levels of H4K16ac and H4K20me3 has been observed.[100] H4K16ac has been implicated in several transcription networks, regulating chromatin structure. Reduction of this modification can perturb these networks, leading to disturbance of transcriptional regulation, and genetic instability.[100] H3K20me3 has been implicated in DNA damage control and a reduction of this modification could increase mutations and lead to genomic instability. Reduction of H3K20me has also been correlated with increased malignancy.[100]
Also shown to be involved in cancer development is the epigenetic silencing of genes. Similar to mutations in tumor suppressors, DNA methylation can prevent the transcription of these genes. Many genes can be deactivated in a single cell through errors in the DNA methylation mechanism, potentially creating an ideal environment for tumor growth.[96] Just as DNA methylation has been shown to participate in cancer development, the involvement of silencing histone modifications in cancer has also been investigated. The first steps towards implicating silencing histone modifications in cancer development have been done on the Polycomb Group protein complex, which was found upregulated in aggressive prostate tumors. It contains histone methyltransferases responsible for the methylation of H3K9 and H3k27, and upregulation could therefore cause increased silencing.[65] It was observed that disturbed balance of H3K27me3 can promote tumor development.[101] Another interesting implication of histone modifications in silencing in cancer cells is the observation that bivalent genes often shift to repression. Several tumor suppressors were shown to gain the repressive marks H3K9me2 and H3K9me3 in addition to their normal bivalent marks. The increase of repressive modifications on these genes tips the scale towards full repression.[65] The addition of repressive marks to repress these genes indicates that this repressive state might be reversible. If the additional repressive marks H3K9me2 and H3K9me3 were to be removed, the bivalent state would return and the tumor suppressors would no longer be repressed. This could provide interesting ways of cancer treatment. When observing cancer cells with repressed bivalent domains in tumor suppressors, a possible treatment could be to switch the bivalent domain to active. Using a cure involving demethylation of H3K9, and possible acetylation to activate the gene, could tip this scale of the bivalent domain, activating the tumor suppressor. Interestingly, one study suggests that several bivalent domains are enriched for DNA Polymerase II.[102] The connections found on promoters between DNA Polymerase II and H3K4 methylation could suggest that initiating transcription through DNA Polymerase II could establish H3K4me presence, or vice versa.
Conclusively, histone modifications patterns are increasingly used in the diagnosis of cancer. They can be used to distinguish between different states of tumor development, and improve our knowledge on the origin of the cancer.

Epigenomic treatments against cancer are being increasingly discovered

As more becomes known about the epigenome in tumor cells, it becomes more evident that this could prove an important field for new cancer treatment. Considering that many epigenomic changes are reversible, developing drugs that target proteins involved in the (de)acetylation or (de)methylation of histones or the DNA could potentially cure these cells. Several drugs targeting epigenetic proteins have already been developed, and many more are in development.

Azanucleoside drugs function as inhibitors of DNA methyltransferase enzymes, and they are already being used in several countries. Other nucleoside inhibitors of DNA methylation, such as Zebularine and 5-fluoro-2'-deoxycytidine, are in different stages of development. Other inhibitors of DNA methylation are also being investigated, such as Pro-cainamide or Tea polyphenols.[96] HDAC inhibitors are likewise being used as cancer treatments, such as the recently approved drug SAHA[96] or Depsipeptide, which is currently being used in clinical trials.[33]

Despite positive results from drugs being used so far, there are several arguments against these drugs. They are often non-specific and could target random genes for deactivation or reactivation. The exact pathway through which these drugs function is unknown and increasing our understanding of this process could increase their efficiency. The occurrence of side-effects due to their wide range of targets is currently unknown and definitely needs to be investigated.[11] For DNA methylation inhibitors it was found that they mainly target dividing cells, focusing their effect on the rapidly dividing cancer cells. These drugs also seem to prefer targeting abnormally silenced DNA over normally silenced regions. What causes this specificity is currently unknown, but this could prove to be valuable information for the development of future DNA methylation protein targeting drugs.[99] Despite our current lack of knowledge on the exact workings of epigenetic drugs, several are working efficiently as cancer treatment, and have been approved for use, indicating a great potential for the use of epigenetic cancer treatments in the future.

Epigenomic patterns also hold potential for the treatment of other diseases

The involvement of epigenomics have been not only been implicated in cancer, but also in several other diseases. In Rubinstein-Taybi Syndrome, which is characterized by mental retardation, seizures and heart defects, mutations deactivate CREBBP. This deregulates the acetylation of histone H3 of important genes such as p53 or TFIIE. In Coffin-Lowry Syndrome a mutation deactivates RSK2, a protein involved in the activation of CREBBP. This causes symptoms similar to Rubinstein-Taybi Syndrome. The dysregulation of CREBBP in both syndromes results in general deregulation of DNA transcription, due to severe deacetylation.[103] Increased knowledge of the pathways of histone (de)acetylation, and the proteins involved, has identified the pathways that cause these syndromes and may eventually lead to the development of a cure.

Epigenetic regulation has also been implicated in many personality and psychiatric disorders. Studies show that many neuropsychiatric genes are repressed in several disorders, such as schizophrenia[104] and borderline personality disorder (BPD).[105] In a study on DNA methylation in BPD patients, several neuropsychiatric genes (such as MAOA, HTR2A and S-COMT) showed significant increase of DNA methylation compared to controls. DNA methylation generally represents repressed genomic regions. These genes therefore seem to be increasingly repressed in BPD patients.

Deregulation of genes and possible related pathways through epigenetics is a perspective that has to be taken into account when researching disorders, because these examples show that not only mutated genes can cause deregulation, and thus expression of the disorder.[105]

Epigenomic patterns have helped to create better models and treatments

Above examples show the importance of epigenetics in the potential development of novel treatments to various diseases. For this it is important to develop better models regarding epigenetic transcriptional regulation. Despite the increase in our knowledge in this throughout the last decades, many questions remain to be answered. For one, despite their importance, we still know little about presence of histone modification patterns on silencers. Some reports speculate about the presence of

certain modifications, as described in chapter 3. This however is only the beginning of the extensive research necessary to grasp the full extent of silencer activity. It might be possible that silencers lack specific histone modification patterns. If this is true, different identification methods will have to be explored.
We also know little about the signals histone modifying proteins receive. It was found that some receive extracellular signals. The communication between the extracellular environment and chromatin remodeling factors holds interesting implications. Extracellular cues can influence many cells at once, potentially causing tissue-wide epigenomic remodeling. There is also relatively little known about the direction histone modifying enzymes receive. How do they target the correct residue on a specific histone? It is considered that short and long RNAs are able to regulate this precise localization, but research to this function of RNA fragments is still preliminary.
Another important topic is that most cells are able to survive deletion of some histone modifying enzymes. Is this caused by redundancy between the different modifications? Or are these just marks and not required for any biological function? This last possibility might seem true for some genomic characteristics in certain organisms, such as H3K4 trimethylation in yeast, but certainly not for all. Expansion of research on this subject could show which modifications are required for survival and which aren't, and if these unnecessary histone modification confer any biological function.

Conclusively, these are just some of the questions that require an answer before we can understand the regulation of transcription of the human genome. This regulation has proven to be so complex, that we might never know its full extent. Identification of epigenomic patterns however has greatly increased our understanding of their localization and function, and continue to do so. It has also increased the development of drugs and treatments for world-wide and deathly diseases. It is our hope that future research on epigenomic patterns will increase the development of new treatments and drugs for these and other diseases.

References

1. Consortium IHGS. (2004) Finishing the euchromatic sequence of the human genome. Nature 431(7011):931-945.
2. J. C. Venter et al. (2001) The sequence of the human genome. Science 291(5507):1304-1351.
3. Z. Wang et al. (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. 40(7):897-903.
4. D. Schübeler et al. (2004) The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes & Dev. 18:1263-1271.
5. F. Thoma, T. H. Koller and A. Klug (1979) Involvement of Histone H1 in the organization of nucleosome and of the salt-dependent superstructures of chromatin. J. Cell Biol. 83(2 pt 1): 403-427.
6. R. D. Kornberg and Y. Lorch (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell 98(3):285–294.
7. R. M. Raisner et al. (2005) Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. Cell 123(2):233–248.
8. J. Ausió (2006) Histone variants - the structure behind the function. Brief Funct. Genomic Proteomic 5(3):228-243.
9. Maze et al. (2014) Every amino acid matters: essential contributions of histone variants to mammalian development and disease. Nat. Rev. Genet. 15(4):259-271.
10. C. Koch et al. (2007) The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res. 17(6)691-707.
11. A. Bannister and T. Kouzarides. (2011) Regulation of chromatin by histone modifications. Cell research 21(3):381-395.
12. G. Orphanides, T. Lagrange and D. Reinberg (1996) The general transcription factors of RNA polymerase II. Genes Dev. 10(21):2657-2683.
13. S. Hahn (2004) Structure and mechanism of the RNA polymerase II transcription machinery. Nat. Struct. Mol. Biol. 11(5):394-403.
14. T. Lee and R. Young (2000) Transcription of eukaryotic protein-coding genes. Annu. Rev. Genet. 34:77–137.
15. J. Veuerizas, S. Kummerfield, S. Teichmann and N. Luscombe (2009) A consensus of human transcription factors: function expression and evolution. Nat. Rev. Genet. 10(4):252-263.
16. S. T. Smale and J. T. Kadonaga (2003) The RNA polymerase II core promoter. Annu. Rev. Biochem. 72:449-479.
17. D. H. Lee et al. (2005) Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. Mol. Cell. Biol. 25(21):9674-9686.
18. I. P. Ioshikhes and M. Q. Zhang (2000) Large-scale human promoter mapping using CpG islands. Nat. Genet. 26(1):61-63.
19. B. E. Eddy, G. S. Borman, W. H. Berkeley and R. D. Young (1961) Tumors induced in hamsters by injection of rhesus monkey kidney cell extracts. Proc. Soc. Exp. Biol. Med. 107:191-197.
20. M. L. Atchison (1988) Enhancers: mechanisms of action and cell specificity. Annu. Rev. Cell. Biol. 4:127-153.
21. L. Lettice et al. (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. 12(14):1725-1735.
22. H. Müller and W. Schaffner (1990) Transcriptional enhancers can act in trans. Trends. Genet. 6(9):300-304.
23. S. Ogbourne and T. M. Antalis (1998) Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. Biochem. J. 331(1):1-14.

24. A. J. Courey and S. Jia (2001) Transcriptional repression: the long and the short of it. Genes Dev. 15(21):2786-2796.
25. M. Capelson and V. Corces (2004) Boundary elements and nuclear organization. Biol. Cell 96(8):617-629.
26. N. Nègre et al. (2010) A comprehensive map of insulator elements for the Drosophila genome. PLoS Genet. 15;6(1):e1000814.
27. N. D. Heintzman et al. (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459(7243):108-112.
28. Q. Li, K. R. Peterson, X. Fang and G. Stamatoyannopoulos (2002) Locus control regions. Blood 100(9):3077-3086.
29. J. C. Eissenberg and A. Shilatifard (2006) Leaving a mark: the many footprints of the elongating RNA Polymerase II. Curr. Opin. Genet. Dev. 16(2):184-190.
30. M. Yun, J. Wu, J. L. Workman and B. Li (2011) Readers of histone modifications. Cell Res. 21(4):564-578.
31. S. D. Taverna et al. (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat. Struct. Mol. Biol. 14(11):1025-1040.
32. O. J. Rando (2012) Combinatorial complexity in chromatin structure and function: revisiting the histone code. Curr. Opin. Genet. Dev. 22(2):148-155.
33. A. Munshi et al. (2009) Histone modifications dictate specific biological readouts. J. Genet. Genom. 36(2):75-88.
34. O. Rando (2007) Global patterns of histone modifications. Curr. Opin. Gen. Dev. 17(2):94-99.
35. A. Barski et al. (2007) High-resolution profiling of histone methylations in the human genome. Cell 129(4):823-87.
36. O. Rando and H. Chang (2009) Genome-wide views of chromatin structure. Annu. Rev. Biochem. 78:245-271.
37. A. Ebert et al. (2006) Histone modifications and the control of heterochromatic gene silencing in Drosophila. Chrom. Res. 14(4):377-392.
38. K. M. McGarvey et al. (2006) Silenced tumor suppressor genes reactivated by DNA demethylation do not return to a fully euchromatic chromatin state. Cancer Res. 66(7):3541-3549.
39. K. Zhang et al. (2005) The Set1 methyltransferase opposes lpl1 aurora kinase functions in chromosome segregation. Cell 122(5):723-734.
40. M. Hodl and K. Basler (2012) Transcription in the absence of histone H3.2 and H3K4 methylation. Curr. Biol. 22(23):2253–2257.
41. S. Nakanishi et l. (2008) A comprehensive librry of histone mutants identifies nucleosomal residues required for H3K4 methylation. Nat. Struct. Mol. Biol. 15(8):881-888.
42. I. Maze, K, Noh, A. Soshnev and C. Allis (2014) Every amino acid matters: essential contributions of histone variants to mammalian development and disease. Nat. Rev. Genet. 15(4):259-271.
43. A. Venkatraman et al. (2014) The histone deacetylase HDAC3 is essential for Purkinje cell function, potentially complicating the use of HDAC inhibitors in SCA1. Hum. Mol. Genet. 23(14):3733-3745.
44. Z. Liu et al. (2010) Jmjd1a demethylase-regulated histone modification is essential for cAMP-response element modulator-regulated gene expression and spermatogenesis. J. Biol. Chem. 285(4):2758-2770.
45. P. Rizzo (2003) Those amazing dinoflagellate chromosomes. Cell Res. 13(4):215-217.
46. N. Macdonald et al. (2005) Molecular basis for the recognition of phosphorylated and phosphoacetylated histone h3 by 14-3-3. Mol. Cell. 20(2):199-211.

47. M. Stucki et al. (2005) MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell 123(7):1213-1226.
48. C. Dhalluin et al. (1999) Structure and ligand of a histone acetyltransferase bromodomain. Nature 399(6735):491-496.
49. J. Kim et al. (2006) Tudor, MBT and chromo domains gauge the degree of lysine methylation. EMBO Rep. 7(4):397-403.
50. A. J. Bannister et al. (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410(6824):120-124.
51. J. Mellor (2006) It takes a PHD to read the histone code. Cell 126(1):22-24.
52. Y. Yang et al. (2010) TDRD3 is an effector molecule for arginine-methylated histone marks. Mol. Cell. 40(6):1016–1023.
53. R. Marqueron et al. (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. Nature 461(7265):762-767.
54. Y. Shen et al. (2012) A map of the cis-regulatory sequences in the mouse genome. Nature 488(7409):116-120.
55. N. Nègre et al. (2011) A cis-regulatory map of the Drosophila genome. Nature 471(7339):527-531.
56. B. E. Bernstein et al. (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 120(2):169-181.
57. ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306(5696):636-640.
58. T. Egelhofer et al. (2011) An assessment of histone-modification antibody quality. Nat. Struct. Mol. Biol. 18(1):91-93.
59. H. M. Petrykowska, C. M. Vockley and L. Elnitski (2008) Detection and characterization of silencers and enhancer-blockers in the greater CFTR locus. Genome Res. 18(8):1238-1246.
60. R. E. Thurman et al. (2012) The accessible chromatin landscape of the human genome. Nature 489(7414):75-82.
61. Gene Ontology Consorium (2008) The Gene Ontology project in 2008. Nucleic Acids Res. 36:D440-444.
62. N. D. Heintzman et al. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. 39(3):311-318.
63. S. Kurdistani and M. Grunstein (2003) Histone acetylation and deacetylation in yeast. Nature 4(4):276-284.
64. D. Schübeler et al. (2004) The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes & Dev. 18:1263-1271.
65. A. Lennartsson and K. Ekwall (2009) Histone modification patterns and epigenetic codes. Biochim. Biophys. Acta 1790(9):860-868.
66. M. J. Carrozza et al. (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 123(4):581-592.
67. D. K. Pokholok et al. (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122(4):517-527.
68. C. Jin et al. (2009) H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat. Genet. 41(8):941-945.
69. D. Zilberman et al. (2008) Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature 456(7218):125-129.
70. A. Shilatifard (2008) Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. Curr. Opin. Cell Biol. 20(3):341-348.

71. N.D. Heintzman et al. (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459(7243):108-112.
72. M. K. Choe et al. (2012) Functional elements demarcated by histone modifications in breast cancer cells. Biochem. Biophys. Res. Commun. 418(3):475-482.
73. N. D. Heintzman and B. Ren (2009) Finding distal regulatory elements in the human genome. Curr. Opin. Genet. Dev. 19(6):541-549.
74. G. E. Zentner and P. C. Scacheri (2012) The chromatin fingerprint of gene enhancer elements. J. Biol. Chem. 287(37):30888-30896.
75. S. Spicuglia and L. Vanhille (2012) Chromatin signatures of active enhancers. Nucleus 3(2):126-31.
76. H. M. Herz et al. (2012) Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes Dev. 26(23):2604-2620.
77. M. K. Ma, C. Heath, A. Hair and A. G. West (2011) Histone crosstalk directed by H2B ubiquitination is required for chromatin boundary integrity. PLoS Genet. 7(7):21002175.
78. B. E. Bernstein et al. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125(2):315-326.
79. C. E. Massie and I. G. Mills (2008) ChIPping away at gene regulation. EMBO Rep. 9(4):337-343.
80. S. K. Rhie et al. (2014) Nucleosome positioning and histone modifications define relationships between regulatory elements and nearby gene expression in breast epithelial cells. BMC Genomics 2;15:331.
81. E. Roldan et al. (2005) Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. Nat. Immunol. 6(1):31-41.
82. J. M. Craig (2005) Heterochromatin - many flavours, common themes. Bioessays 27(1):17-28.
83. J. P. Jackson et al. (2004) Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. Chromosoma 112(6):308-315.
84. M. Grunstein (1998) Yeast heterochromatin: regulation of its assembly and inheritance by histones. Cell 93(3):325-328.
85. Y. Tang et al. (2012) Widespread existence of cytosine methylation in yeast DNA measured by gs chromatography/mass spectrometry. Anal. Chem. 84(16):7249-7255.
86. L. Chen and J. Widom (2005) Mechanism of transcriptional silencing in yeast. Cell 120(1):37-48.
87. H. Wang et al. (2004) Role of histone H2A ubiquitination in Polycomb silencing. Nature 431:98-878.
88. J. A. van der Knaap et al. (2005) GMP Synthetase stimulates histone H2B deubiquitylation by the epigenetic silencer USP7. Molecular Cell 17(5):695-707.
89. B. Lehnertz et al. (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at Pericentric heterochromatin. Curr. Biol. 13(14):1192-1200.
90. K. M. McGarvey et al. (2006) Silenced tumor suppressor genes reactivated by DNA demethylation do not return to a fully euchromatic chromatin state. Cancer Res. 66(7):3541-3549.
91. E. Bártová et al. (2008) Histone modifications and nuclear architecture: a review. J. Histochem. Chytochem. 56(8):711-721.
92. S. B. Rothbart et al. (2012) Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat. Struct. Mol. Biol. 19(11):1155-160.
93. B. Lehnertz et al. (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr. Biol. 13(14):1192-1200.

94. I. taniuchi, M. J. Sunshine, R. Festenstein and D. R. Littman (2002) Evidence for distinct CD4 silencer functions at different stages of thymocyte differentiation. Mol. Cell 10(5):1083-1096.
95. Z. Wang et al. (2004) Systematic identification and analysis of exonic splicing silencers. Cell 119(6):831-845.
96. P. Jones and S. Baylin (2007) The epigenome of cancer. Cell 128(4):683-692.
97. J. K. Wiencke, S. Zheng, Z. Morrison and R. F. Yeh (2008) Differentially expressed genes are marked by histone 3 lysine 9 trimethylation in human cancer cells. Oncogene 27(17):2412-2421.
98. F. Barlési et al. (2007) Global histone modifications predict prognosis of resected non small-cell lung cancer. J. Clin. Oncol. 25(28):4358-4364.
99. D. B. Seligson et al. (2005) Global histone modification patterns predict risk of prostate cancer reoccurrence. Nature 435(7046):1262-1266.
100. M. F. Fraga et al. (2005) Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat. Genet. 37(4):391-400.
101. G. van Haaften et al. (2010) Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nat. Genet. 41(5):521-523.
102. J. K. Stock et al. (2007) Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat. Cell Biol. 9(12):1428–1435.
103. D. Waggoner (2007) Mechanisms of disease: epigenesis. Semin. Pediatr. Neurol. 14(1):7-14.
104. H. M. Abdolmaleky et al. (2006) Hypomethylation of MB-COMT promoter is a major risk factor for schizophrenia and bipolar disorder. Hum. Mol. Genet. 15(21):3132-145.
105. G. Dammann et al. (2011) Increased DNA methylation of neuropsychiatric genes occurs in borderline personality disorder. Epigenetics 6(12):1454-1462.

Supplemental 1. Functions of Histone Modifications

Many residues in the core histones and their variants are known to be modified post-translationally. Here, each modified amino acid on histones is described together with the specific modification and the known function of this modification. These modifications are created by specific enzymes, called writers, and can be erased by other proteins, the erasers. For each modification at least one writer has been identified and for some the erasers are also known. The modifications are read and interpreted by proteins called readers, which use specific domains to bind the modification.

	Residue	Modification	Functions	Writer	Eraser	Reader^{Domain}
Histone H1	Lys26	Methylation	Transcriptional silencing	E2h2		
	Ser27	Phosphorylation	Transcriptional activation, chromatin decondenstation			
Histone 2A	Ser1	Phosphorylation	Mitosis, Chromatin assembly, Transcriptional repression	PKC, MSK1		
	Lys4	Acetylation	Transcriptional activation	ESA1		
		Acetylation	Transcriptional activation	CBP, p300, HAT1, TIP60		
	Lys5	Acetylation	Transcriptional activation	Tip60, p300/CBP		
	Lys7	Acetylation	Transcriptional activation	ESA1, HAT1		
	Lys119	Ubiquitylation	Spermatogenesis	RNF1, RING2		
	Thr120	Phosphorylation	Mitosis	NHK1		
	Ser122	Phosphorylation	DNA repair			
	Lys126	Sumoylation	Transcriptional repression	UBC9		
	Ser129	Phosphorylation	DNA repair	MEC1, TEL1		
Histone 2A.X	Ser139	Phosphorylation	DNA repair, Apoptosis	ATM, ATR, DNA-PK	PP4	MDC1^{BRCT}, NBS1^{BRCT}, 53BP1^{BRCT,TDRr}, BRCA1^{BRCT}
	Tyr142	Phosphorylation	Regulation of DNA damage foci formation	BAZ1B	EYA	APBB1^{PID}
Histone 2B	Lys5	Acetylation	Transcriptional activation	p300, ATF2		
	Ser10	Phosphorylation	Apoptosis	STE20		
	Lys11	Acetylation	Transcriptional activation	GCN5		
	Lys12	Acetylation	Transcriptional activity	CBP, p300, ATF2		
	Ser14	Phosphorylation	Apoptosis, DNA repair	MST1		
	Lys15	Acetylation	Transcriptional activation	CBP, p300, ATF2		

	Lys16	Acetylation	Transcriptional activation	GCN5, ESA1		
	Lys20	Acetylation	Transcriptional activation	CBP, p300		
	Lys120	Ubiquitylation	Elongation, Meiosis	RNF20, UBCH6		
	Lys123	Ubiquitylation	Euchromatin	RAD6		
Histone H3	Arg2	Methylation	Transcriptional activation	CARM1, PRMT6	JMJD6	
	Thr3	Phosphorylation	Mitosis	Haspin, Vrk1		Survivin^{BIR}
	Lys4	Acetylation	Transcriptional activation	ESA1, HPA2		
	Lys4	Methylation	Euchromatin, transcriptional activation	MLL, SET1, ASH1, SET7, SMYD2, MLL, ALL-1		CHD^{CHR}, ING Family^{PHD}, RAG2^{PHD}, TAF3^{PHD}, BPTF^{PHD,BRD}, BHC80^{PHD} DNMT3L^{PHD}, PYGO1^{PHD}, JMJD2A^{PHD,TDR}, WDR5^{WD}
		Phosphorylation	Transcriptional activation	PKCβ		
	Arg8	Methylation	Transcriptional repression	PRMT5		
	Lys9	Acetylation	Histone deposition, Transcriptional activation	GCN5, PCAF, SRC-1	SIRT6	BRD4^{BRD}, BAZ1B^{PHD,BRD}
	Lys9	Methylation	Transcriptional silencing, Heterochromatin, DNA methylation	SUV39H1/2, G9a, CLL8, SETDB1, EuHMT1, Riz1, CLR4	JMJD1A/KDM3A, JMJD1B/KDM3B, JMJD1C/TRIP8, JMJD2A/KDM4A, JMJD2B/KDM4B, JMJD2C/KDM4C, JMJD2D/KDM4D	L3MBTL1/L2^{MBT}, HP1^{CHR}, MPP8^{CHR,ANK}, CDY family^{CHR}, TDRD7^{TDR}, UHRF^{PHD}, EED^{WD}, GLP^{ANK}
	Ser10	Phosphorylation	Mitosis, Meiosis, Immediate early gene activation, Transcriptional activation	Aurora B, MSK1/2, SNF1, IKK-α	PP1	14-3-3^{14}
	Thr11	Phosphorylation	Mitosis, DNA damage induced transcription	DLK		
	Lys14	Acetylation	Transcriptional activation, Elongation, DNA repair, RNA Pol II transcription, RNA Pol III transcription, Euchromatin	GCN5, PCAF, CBP, p300, MOZ, MORF, TIP60, SRC-1, ESAL, TIP60, ELP3, HPA2, TAF1, SAS2/3		BRD4^{BRD}, BAZ1B^{PHD,BRD}, BRG1^{BRD}
	Arg17	Methylation	Transcriptional activation	CARM1		TDRD3^{TDR}

	Lys18	Acetylation	Transcriptional activation, DNA repair, DNA repliction	GCN5, PCAF, CBP, p300		
	Lys23	Acetylation	Histone deposition, Transcriptional activation, Elongation, DNA repair	GCN5, PCAF, p300, SAS3		
	Arg26	Methylation	Transcriptional activation	CARM1		
	Lys27	Acetylation	Transcriptional activation	GCN5		
	Lys27	Methylation	Transcriptional silencing, X Inactivation	EZH2, EZH1, G9a	JMJD1A/KDM3A, JMJD1B/KDM3B, KDM6A/UTX, JMJD3/KDM6B	Pc[CHR], CDY Family[CHR]
	Ser28	Phosphorylation	Mitosis, Immediate-early gene activation	Aurora B, MSK1/2		14-3-3[14]
	Lys36	Acetylation	Transcriptional activation, Elongation	GCN5, PCAF, SET2		
	Lys36	Methylation	Transcriptional elongation	NSD1, SET2, SMYD2, NSD2	JHDM1A/KDM2A, JHDM1B/DM2B, JHDM3A/KDM4A, JHM3B/KDM4B, JHDM3C/KDM4C, JHDM3D/KDM4D	MRG5[CHR]
	Thr45	Phosphorylation	DNA replication, apoptosis	PKCd		
	Lys56	Acetylation	DNA damage repair, Chromatin assembly, Transcriptional activation	CBP, p300, SPT10		
	Lys79	Methylation	Transcriptional activation, Elongation, Euchromatin, Checkpoint response	DOT1		
Histone H4	Ser1	Phosphorylation	Transcriptional activation, DNA repair, Mitosis, Chromatin assembly	CKII		
	Arg3	Methylation	Transcriptional activation, Transcriptional repression	PRMT1, PRMT5	JMJD6	TDRD3[TDR]
	Lys5	Acetylation	Histone deposition, transcriptional activation, DNA repair	ATF2, HAT1, CBP, p300, TIP60, HBO1, ESAL, TIP60, HPA2		BRD4[BRD]

Lys8	Acetylation	Transcriptional activation, DNA repair	CBP, p300, TIP60, HBO1, GCN5, PCAF, ESAL, TIP60, ATF2, ELP3		
Lys12	Acetylation	Histone deposition, Transcriptional activation, Telomeric silencing, DNA repair	HAT1, CBP, p300, TIP60, HBO1, ESAL, TIP60, HPA2		$BRD2^{RD}$, $BRD4^{BD}$
Lys16	Acetylation	Transcriptional activation, DNA repair, Euchromatin	MOF, TIP60, GCN5, ESAL, TIP60, ATF2, SAS2	SIRT1	
Lys20	Methylation	Transcriptional silencing, Heterochromatin	PR-SET7(mono), SUV420H1(di), MMSET, SUV420H2(tri)		$L3MBTL1^{MBT}$, $MBTD1^{MBT}$, $JMJD2A^{PHD}$, $PHF20^{PHD}$, $53jBP1^{BRCT,TDR}$
Lys59	Methylation	Transcriptional silencing			
Lys91	Acetylation	Histone deposition, DNA damage repair, Chromatin assembly	HAT1		

Domain Key

14	14-3-3 domain	BRD	Bromodomain	PHD	PHD zinc finger domain
ANK	Ankyrin repeat domain	CHR	Chromodomain	TDR	Tudor domain
BIR	Baculovirus IAP repeat	MBT	MBT domain	WD	WD40 repeat domain
BRCT	BRCA1 C-term domain	PID	Phosphotyrosine interaction domain		

Layman Summary

The human body contains billions of cells, collected in organs, like the hearth or the lungs, performing specific functions. The cells in these organs also have their specific function to be carried out, and they have corresponding features. In order to contain and maintain both these features and their function, they possess a specific set of proteins. Proteins are molecules that perform all functions of cells, and thus the human body. These proteins are transcribed from the DNA, the blueprint of the human body that is present in each cell. Every cell contains the entire DNA, but only several proteins are required in a cell. This is why most parts of the DNA are silenced, preventing the transcription of the unwanted proteins in a specific cell. This silenced DNA is called heterochromatin and the DNA that is transcribed is called euchromatin.

The DNA is wrapped around a protein complex, consisting of histones. Four different histones form this complex (Histone H2A, H2B, H3 and H4) and all of them have protein tails protruding outward from the complex. These tails can be modified by the addition of specific molecular groups, such as methyl or acetyl. These modifications are found on most histones, although not all modifications are found on every histone. These histone modifications can form specific patterns along the DNA. These patterns are being annotated to several genomic regulatory elements. Genomic regulatory elements are parts of the DNA that enable regulation of the transcription of proteins. There are proximal and distal regulatory elements. Proximal regulatory elements are the promoters. These contain the transcription start site, where the transcription of a protein is initiated. Distal regulatory elements can either enhance (enhancer) or reduce (silencer) the transcription of these proteins, often by direct interaction with the promoter or by using different transcription factors. Transcription factors are proteins that can stimulate (activator) or diminish (repressor) the transcription by binding to the transcription complex or the promoter or distal regulatory elements.

The histone modification patterns that have been discovered and annotated to specific regulatory elements have been used to identify more of these elements along the DNA. Through this many enhancers and promoters in different cell lines have been discovered. These patterns can also help discover new patterns in diseased cells, such as cancer cells. It is possible that in these diseased cells certain regulatory elements are non-functional. Comparing the histone modification patterns with healthy cells can help identify these non-function elements. This could provide great new opportunities for diagnosis and treatment. Other diseases than cancer are also affected by histone modifications, and potentially also benefit from increased knowledge on these modifications and their patterns.

Unfortunately, not all regulatory elements have identified histone modification patterns. Silencer, for example, have not been researched much and no patterns are known. In this work we provide insight in the histone modifications and their patterns of the different regulatory elements, and we specify what is known about histone modifications in silencers. Functional applications and social relevance of this knowledge is also discussed.